An Application of Eyegaze Tracking for Designing Radiologists' Workstations: Insights for Comparative Visual Search Tasks

M. STELLA ATKINS and ADRIAN MOISE
Simon Fraser University
and
ROBERT ROHLING
University of British, Columbia

The goal of this research is to use eyegaze tracking data to provide insights into designing radiology workstations. We designed a look-alike radiology task with artificial stimuli. The task involved a comparative visual search of two side-by-side images, using two different interaction techniques. We tracked the eyegaze of four radiologists while they performed the task and measured the duration of the fixations on the controls, the left and right images, and on the artificial targets. Response time differences between the two interaction techniques exceeded the differences of fixations on the controls. Fixations on the left-side images are longer than the right-side images, and the search for multifeatured targets occurs in two phases: first a regular scan path search phase for a likely target and then a confirmation phase of several fixations on the target in each side-by-side image. We conclude that eyegaze tracking shows that disruption of visual search leads to cognitive disruption; subjects use the left image as a reference image and multiple saccades between left and right side images are necessary, because of the limitations of the visual working memory.

Categories and Subject Descriptors: H.1.2 [**Information Systems**]: Models and Principles-User/Machine Systems—*Human factors; human information processing; software psychology design*; H.5.2 [**Information Interfaces and Presentation**]: User Interface—*Evaluation/methodology; ergonomics; theory and methods*; I.4.8 [**Computing Methodologies**]: Image Processing and Computer Vision—*Scene analysis; tracking*; J.3 [**Computer Applications**]: Life and Medical Sciences

General Terms: Design, Human Factors, Measurement

Additional Key Words and Phrases: User interface evaluation, eyegaze tracking, comparative visual search, human-computer interaction, cognition, radiologist productivity

---

1. INTRODUCTION

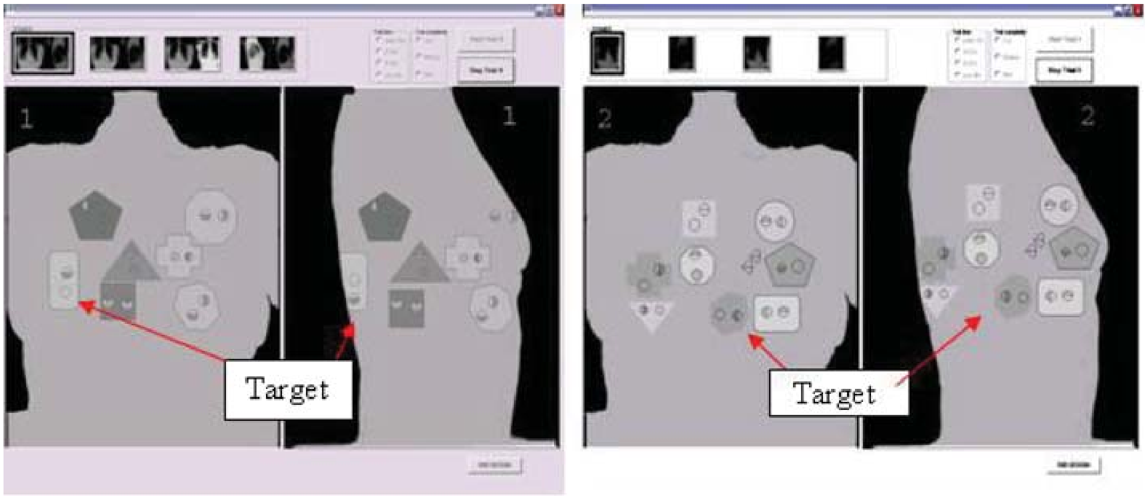
1.1 Overview

Our research focuses on designing new interaction techniques to improve radiology workstations, which are used by radiologists to view patient images to make diagnoses. Radiologists perform diagnostic tasks by interpreting details on patient images. The diagnostic task has stringent requirements of accuracy

---

Authors' addresses: M. S. Atkins and A. Moise, School of Computing Science, Simon Fraser University, Burnaby, B.C. V5A 1S6, Canada; email: {stella,amoise}@cs.sfu.ca; R. Rohling, Department of Electrical and Mechanical Engineering, University of British Columbia, Vancouver, B.C. V6T 1Z4, Canada; email: rohling@ece.ubc.ca.

ACM Transactions on Applied Perception, Vol. 3, No. 2, April 2006, Pages 136–151.



(a) Stages interaction technique

(b) FUI interaction technique

Fig. 1. Screen layout for the task under (a) Stages and (b) FUI interaction techniques. The images to be displayed are selected by clicking on the icons in the top left. Both images must be viewed to detect a target. (a) Study 1; (b) study 2.

to prevent errors, yet it is often repetitive; speed is also necessary to complete the workload in a timely way. A goal of radiology workstation design is to develop workstation interaction techniques to help the radiologists perform their tasks efficiently [Moise and Atkins 2004b]. In previous studies, we have shown how sophisticated hanging protocols called Stages, which automate the image organization and display, can help speed up the interpretation process [Moise 2003; Moise and Atkins 2002, 2005]. In traditional image interfaces called free user interface (FUI), the images are selected from thumbnails of individual image sequences, so for comparison studies involving two or more images, the display has to be selected and chosen using multiple clicks on individual thumbnails. However, in our new Stages interaction technique, images are selected and presented as image pairs corresponding to the radiological workflow for comparing images, so one click can display a new image pair for comparison. In more recent work, we showed that students performing a comparative visual search task for artificial targets had similar response times, mouse clicks, and errors to radiologists performing the same task [Moise et al. 2005] using the same interaction technique—either Stages or FUI. There were major differences between two interaction techniques. For both subject groups, subjects used significantly fewer mouse clicks in Stages and were significantly faster using Stages than using FUI.

We justify using artificial targets instead of nonradiology images for the sake of gaining experimental control. These controlled, abstract conditions permit accurate estimation of the time and errors when performing the tasks with each interaction technique.

Yang et al. [2002] proposed using eyegaze tracking to solve many applications problems in visual search tasks. We show here that eyegaze tracking is a useful tool to help design and assess workstation interaction techniques. We also use eyegaze patterns to examine the use of visual memory in visual perception during the search for a matching object (a target) in side-by-side images, and examine the impact of this on the response time and on errors.

To introduce both our radiology look-alike task and the data obtained from eyegaze tracking, typical screen layouts for the task are shown in Figure 1.

Figure 1a shows the layout for the Stages interaction technique and (b) shows the layout for the FUI interaction technique. Both layouts consist of a left and right viewport containing the stimuli images,

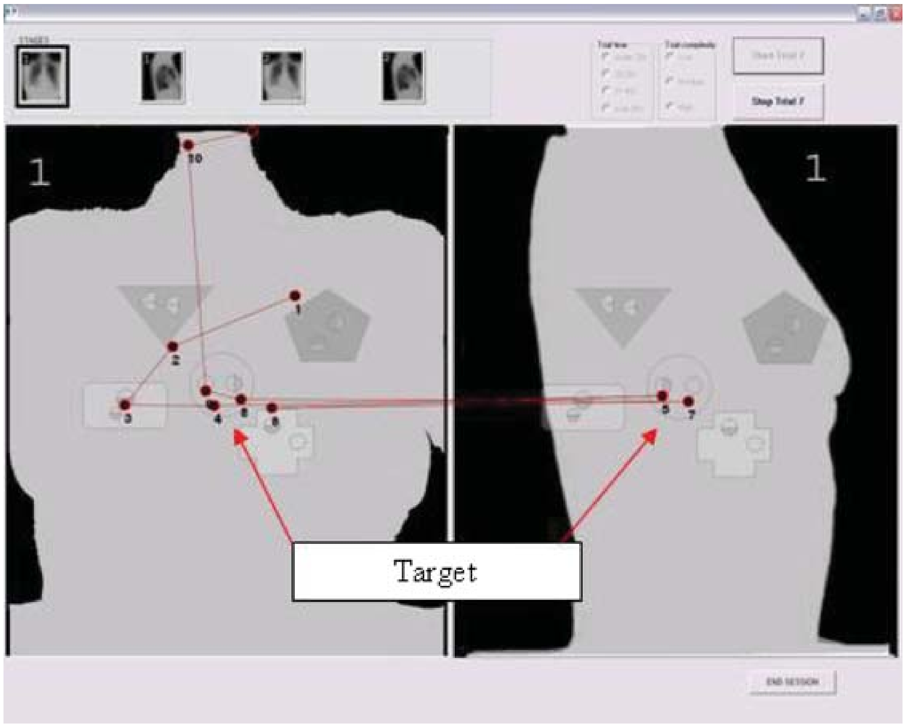


Fig. 2. Typical eyegaze pattern for a trial under the FUI interaction technique (Subject 2, trial 7, study 1, FUI). The numbered dots represent fixations in temporal order and the lines between the dots represent the saccades between the fixations.

the controls used to start/stop the current trial for image selection (upper right), and image navigation controls (upper left; these controls are the only difference between the Stages and FUI techniques). In both techniques, the images to be displayed are selected by clicking on the icons in the controls at the top left. Two images must be viewed to detect a target (a pair of matching split discs). Details of the techniques and targets are given in the methods section. Each trial has two studies, labeled 1 and 2 in the upper corner of the images. If there is a target in study 1, there may or may not be a target in study 2. If there is a target in both studies, the target will be in the same corresponding location. If there is no target in study 1, there may or may not be a target in study 2. There is, at most, one target in a study (a pair of images). These situations correspond to radiologists searching for tumors in current and in past studies, and detecting whether or not the tumor has changed in size.

Figure 2 shows a typical eyegaze pattern during study 1 of a trial under the FUI interaction technique. The raw eyegaze data is obtained and analyzed into a sequence of fixations, as described in the methods section.

The numbered dots correspond to the eyegaze fixations of duration $>$100 msec in temporal order and the lines between the dots show the saccades between the fixations. From the locations and durations of the fixations, we can obtain information on the time spent viewing the workstation controls, the stimuli images, and the dynamic patterns of the eyegaze fixations.

1.2 Motivation for Eyegaze Analysis

1.2.1 *Explain Performance Effects.* In previous work we found that the average response time for the Stages interaction technique was significantly less than for the FUI technique, showing a 9–14%

reduction in response time [Moise and Atkins 2004a; Moise et al. 2005]. The goal of the eye movement analysis is to show how visual attention is allocated over the controls and the targets as a function of the Stages and FUI conditions, in order to explain these timing differences.

1.2.2 *Explain Decision Processes.* The response time is affected by the trial outcome, because of the nature of our particular experimental design, which includes, at most, one target per study. Trials that have a target will be faster, because the subject can stop the trial once a target has been found. However, trials with no targets should have higher response times, because the subject has to perform a full search over all the objects. Eyegaze studies should reveal how target decisions are made, including searching and recognition phases as others have reported [Pomplun et al. 2001; Inamdar and Pumplun 2003; Kundel 2004]. The first phase, called the *search phase*, occurs while the subject is searching the whole image for a suspicious object, or a likely target. The second is called the *recognition phase*, when multiple fixations are used to confirm the target.

1.3 Related Work on Perception and Visual Search

1.3.1 *Perception in Radiology.* Although there is a huge body of research on perception during reading, there is much less on visual search [Rayner 1998], and even less on radiology, where filmless diagnosis using computer monitors has only become more common in the last decade. Radiology training is designed to provide assumptions about normal and abnormal medical images. However, radiologists do make mistakes. General estimates suggest that, overall, there is a 20–30% miss or false negative rate in radiology, with a 2–15% false positive rate, a 10–20% interobserver variation, and 5–10% intraobserver variation [Hendee 2002]. Many lesions can be seen by peripheral vision and verified by foveal vision, but peripherally inconspicuous targets must be found by scanning the fovea over the image. The resultant search pattern, as shown by eye-position recording, is influenced by both the patient's clinical history and the radiologist's experience [Kundel et al. 1978]. Their model assumes that an eye fixation picks up a certain amount of visual information from foveal and peripheral inputs. Thus, in order to apply this idea to analysis of scanning fixations during search, Kundel et al. invented fixation clusters and assumed a 5° visual field for information pick up. The boundary of a cluster of fixations extends radially 2.5° from the centroid of the cluster and can be regarded as the center of focal attention [Nodine et al. 1992]. The assumption here was that the subject was focusing on some aspect of the display such as the target, as described in [Nodine et al. 2001a]. This led to a 5^o visual field to define regions of interest (ROIs) in images.

False-positives can be explained by image noise and overlapping anatomic structures that often mimic disease entities. However, false-negative errors are harder to understand, especially when missed lesions can be seen in retrospect [Krupinski 2000]. Based on gaze fixation time, Kundel and colleagues [Kundel et al. 1978; Nodine and Kundel 1987] have proposed three classifications for false-negative errors. These errors can be classified as (1) faulty visual search or failure to fixate the abnormal region, (2) faulty pattern recognition or failure to report a lesion that has been fixated for a short amount of time, and (3) faulty decision making or failure to report a lesion that has been extensively fixated [Kundel 2004]. Berbaum et al. [2001] reports that the results of numerous studies have indicated that the temporal threshold separating recognition and decision errors should be about 1.0 s of dwell time for pulmonary nodules [Kundel et al. 1978, 1987; Nodine and Kundel 1987], skeletal fractures [Hu et al. 1994], and diverse chest abnormalities [Berbaum et al. 1998]. However, more recent work in mammography has used a shorter fixation time of only 200 ms to separate recognition and decision errors [Manning et al. 2004]. In our research, we used the fixation-duration histogram data to determine a separation of errors of recognition from errors of decision making.

The noise in medical images limits the perception of contrast and detail. Current research is extending the mathematical models that relate signal detection to physical descriptions of image signal and noise to include realistic lesions and anatomical backgrounds. Recent research is focusing on issues of conspicuity, which may be used to determine the boundary between recognition and decision errors [Krupinski et al. 2004; Kundel and Nodine 2004; Mello-Thoms et al. 2002; Manning and Ethell 2002], but the models currently available are task-specific for mammography or for lung nodule detection, and are not suitable for general use.

Perception research can benefit clinical radiology by suggesting methods for reducing observer error, providing objective standards for image quality, and providing a scientific basis for image-technology evaluation. The Medical Image Perception Society was created in 1997 to promote research in several aspects of modeling and evaluating computer-aided perception tools in radiology [Krupinski et al. 1998; MIPS 2005].

1.3.2 *Eyegaze Patterns During Comparative Visual Search.* Only a few other researchers, notably Pomplun [Pomplun et al. 2001a, 2001b; Pomplun and Ritter 1999], have studied eyegaze patterns during comparative visual search tasks, although picture-matching tasks involving recognition of the same object within two pictures have been much studied [Humphrey and Lupker 1993]. Like other researchers such as Kundel [2004] have reported, Pomplun et al. found that there are two phases of visual search—a first phase of search and comparison, and a second phase of detection and verification. The limitations of the working visual memory require that once a suspicious object has been seen, a second phase of verification is required. In general, their results suggest that oculomotor behavior in comparative visual search is determined on the basis of *cognitive economy* and that extra saccades are preferred to extra effort to remember multiple features in dense clusters across saccades. These researchers have been proposing answers to questions, such as, how are objects represented in memory and how are objects recognized, whereas our research focuses on how an application design can benefit from the results of eyegaze tracking.

1.4 Workstation Ergonomics

Eyegaze studies have an increasingly important role in examining usability of human–computer interfaces [Duchowski 2000]. We are concerned not just with the contrast and resolution of abnormalities, but also with the ergonomics of the entire workstation [Moise and Atkins 2002; Moise and Atkins 2004b]. Using the workstation controls causes an interruption of the visual search, and the importance of keeping eyegaze fixations on the medical images to minimize visual distraction cannot be overstated. In a study to compare film and CRT viewing of radiological images, Krupinski and Lund showed that radiologists spend up to 20% of fixation clusters on the workstation controls menu [Krupinski and Lund 1997]. In a recent discussion in the Society for Computer Applications in Radiology (SCAR) Newsletter on the pros and cons of automated speech recognition in radiology workstations, the distraction of viewing and correcting the speech recognition text during the radiological diagnostic task was cited as a major deterrent to adopting automated speech recognition in radiological workstations [SCAR 2004].

Our research focuses on improving the workstation controls in order to allow the radiologist to focus on the recognition and decision perception processes, while minimizing the interruption of the visual search.

2. MATERIALS AND METHODS

2.1 Participants and Environment

The subjects were four senior radiology fellows with normal or corrected to normal vision. All the experiments took place in a room at the hospital with a significant level of nearby traffic and with

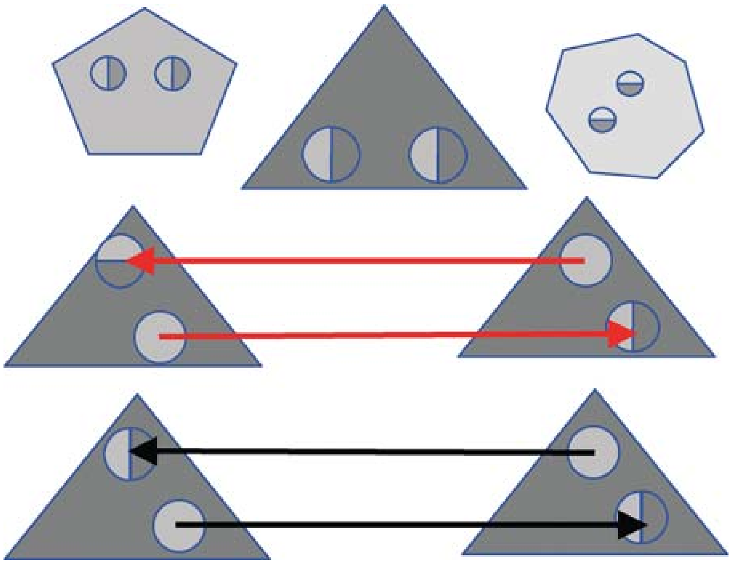


Fig. 3. *Top row*: typical targets of 2 spherical disks of the same size split in half in the same direction, either vertically or horizontally. *Middle row*: A potential target seen in the comparative search over two images: the potential target is incompletely presented on each image, instantiated with the wildcard shown as the source of the arrow. However, the disk is split horizontally (left hemifield image) and vertically (right hemifield image), so it is a distracter. *Bottom row*: The wild card instantiates in a disc with the same divider orientation, so it is a target.

a noticeable background noise. While these conditions were not ideal, it did not seem to bother the radiologists—we assume they are quite used to focus and work in similar conditions. We also videotaped the subjects performing the task.

2.2 Targets and Stimuli

We used abstract targets and images rather than real radiology images; however, we designed our stimuli to reflect the radiologists task, using greyscale images. The target is an item with two discs of the same size, half-split along the same vertical or horizontal diameter, and half-shaded. Three examples of targets are presented in the top row of Figure 3. Images also contained distracters, taking forms such as unequally sized discs, hearts, or octagonal-sided discs. We presented the targets in such a way a subject must discriminate a target from a distracter solely by integrating the information from two related images displayed on a right and left viewport (i.e., the target is incompletely revealed to the user because of partial occlusion). The occlusion is simulated in our stimuli with the introduction of a "wild-card," which forces our subjects to register information between the two images of a study, in a comparative visual search. A wild card is a disc with a uniform hollow fill and is used to represent the disc divider, as seen in the middle row of Figure 3. A disc with a uniform fill can hide a disc that could be divided either vertically or horizontally. The user must find on a related image the actual instantiation of a wild card. Depending on the orientation of the occluded disc divider, a wild card could either instantiate into a distracter (middle row), or into a target (bottom row) in Figure 3. Registration is required for solving the "wild card" into a target. Only the orientation of the divider is important. It does not matter which half of the disc (top or bottom for a horizontal divider, and, respectively, left or right for a vertical divider) is grayed-out. Another situation, also corresponding to a distracter, occurs when the wild card does not instantiate into a divided disc, but remains as another hollow disk. Note a potential target always has a wild card. Consequently, for every potential target containing a wild card, a subject has to register complementary information from the two images of the same study.

A similar occlusion occurs frequently in radiology because of anatomical structures shown as bright areas in the image, which overlay features of interest. Such is the case of a barely visible lung nodule hidden behind a rib on a chest CR, or a liver tumor hidden behind a blood vessel.

To simulate the radiologist's follow-up on a radiographic examination, we introduced a time dimension by presenting to our subjects two instances of the same scene, corresponding to different time moments. Hence we asked our subjects to detect the target from the two images in study 1 and then track the evolution of the target in time. Therefore, each trial consisted of two studies, where each study has two images. The two images of study 1 were presented first and the two images of study 2 had to be viewed next, to detect the evolution in size of any target seen in study 1.

2.3 Eyegaze Tracking Equipment

The eyegaze tracking system used is an Applied Science Laboratories (ASL) Model 504, where the stimulus presented to the subject is restricted to a planar surface, such as a computer monitor, and where head mounted optics are not desirable [ASL 2000]. The system allows the subject approximately one square foot of head movement that eliminates the need for head restraint; however, greater accuracy is possible if the subject is using a chin rest, as head movements by moving toward the monitor will cause inaccuracies. To display the stimuli images we used a 17″ Samsung 770 TFT LCD monitor, with a resolution of 1280×1024, brightness of 220 cd/m^2, contrast ratio of 400:1, and viewing angle of 160/160 (horizontal/vertical). The subjects sat 55 cm from the screen with their chin in a chin rest.

The system uses bright pupil technology, giving good capture and contrast. The system emits near-infrared light invisible to the human eye, to create bright pupil and cornea reflection signals. Since the system measures the distance between pupil center and the cornea reflection, it is sensitive to rotation, but not translation. The data were sent to a PC through a serial connection and saved to an ASCII log file. Each eyegaze sample was averaged over four fields using the ASL control software to smooth out small-magnitude eyegaze position jitter. The eye-tracker control software ran on an IBM 390E laptop. Focus Enhancements TView Gold 300 scan converter was used to create a composite video frame of eyegaze location overlayed on the scene (i.e., the frames captured of the experimental task). The experimental task was implemented as a multithreaded Visual C++ application, to separate the user interaction thread from the data recording thread. Recorded data include time, x and y eye-position coordinates and pupil diameter, with a frequency of 60 Hz, an accuracy of 0.5° visual angle, and a resolution of 0.25° visual angle. Eye position coordinates correlate to specific areas on the screen being viewed after a calibration procedure.

2.4 Eyegaze Fixations

Fixations are assumed to be of at least 100 ms duration and are calculated from the points of gaze using a dispersion threshold algorithm based on an algorithm developed by [Salvucci and Goldberg 2000]. For our analysis, we defined a fixation cluster to subtend an angle of 3° at the fovea, which is about the size of our image features, such as targets and distracters. Although this is smaller than the 5° used by Kundel [Kundel et al. 1978], we found this was large enough to detect all the fixations and gave the fixation centroid on appropriate image features.

2.5 Experimental Procedure

The screen layout for the task, shown in Figure 1, consists of left and right viewports containing the stimuli images, the controls used for image selection, and the controls used to start/stop the current trial. Subjects were asked to find an abstract target on a grey background. Each subject performed two consecutive blocks of 15 trials, one for each of the two different interaction techniques. In each trial, a target (if present) had to be located in the first study set of two images, and its evolution noted in a second study set of two images. The subjects were asked to identify where they found a target by pointing with the mouse and saying "here is the target." This was recorded on camera and used to decide which trials were correct and which trials had an error. Each subject performed the same 30 trials; two started with stimuli set A and the other two started with set B.

Instructions about the task were given using several training steps presented on the computer screen. Each training step was followed by a short practice session, where the subjects' understanding of the recently learned concepts was tested, before the trials were started.

2.6 Interaction Techniques

The software allowed for the simultaneous display of only two images. The subjects had to interact with the system to see the two images from the first study and then the two images from the second study, and, in case a target was present in both studies, make a comparison in size of the target between one image from the first study and one image from the second study. Two different interfaces were provided for user interaction (Stages and FUI), each using four thumbnail-size controls located in the top-left side of the screen. With the Stages technique, only one mouse click on the controls (the second icon from the left) is required to load the two images from study 2 into the viewports; comparisons between present and past views could be made by clicking on the third and fourth icons. With the FUI technique, users had to first select a viewport, then select the image icon to load into the selected viewport. Hence four mouse clicks (two for selecting viewports and two for selecting images) were needed to load the two images for study 2. Comparisons between the same views for present and past images could be made by loading the appropriate images from the thumbnail icons into the selected viewports.

2.7 Independent Variables

We had three independent variables: the interaction technique, the trial complexity, and the trial outcome. There are two conditions for the interaction techniques (Stages and FUI). For the trial complexity, we used low, medium, or high complexity stimuli images, depending on the number of distracters and potential targets [Moise 2003; Moise et al. 2005b]. We used the following notation convention for trial outcome: "0" means no target present in the study, "1" means a target was present. We distinguish the following trial outcome conditions: "00," "01," "10," "11 same," and "11 diff" where the terms "same" and "diff" refer to the size of the target and whether it is the same in both studies or different. Consequently, our experimental design has $2 \times 3 \times 5 = 30$ distinct combinations. We associated one trial per combination (where a trial is a set of two studies, where each study has two images), for a total of 30 trials. The stimuli for the 30 trials were grouped into two disjoint blocks of 15 trials, one for each interaction technique.

2.8 Dependent Measures

The following were the dependent measures used.

1. Response time. Individual trial completion times were recorded in a user-specific log file. The response time was measured from the moment the subject clicked on "Start trial" to have the stimuli displayed, until the subject clicked on the "Stop trial."
2. Interpretation errors. The accuracy of each trial was assessed by video analysis. The video footage captured both the screen content and the subject's verbal interpretation.
3. Eyegaze data. Eyegaze measures, including the foveal fixation position on the monitor, were collected 60 times per second. Eyegaze measures included the gaze time on the stimuli and on the applications'

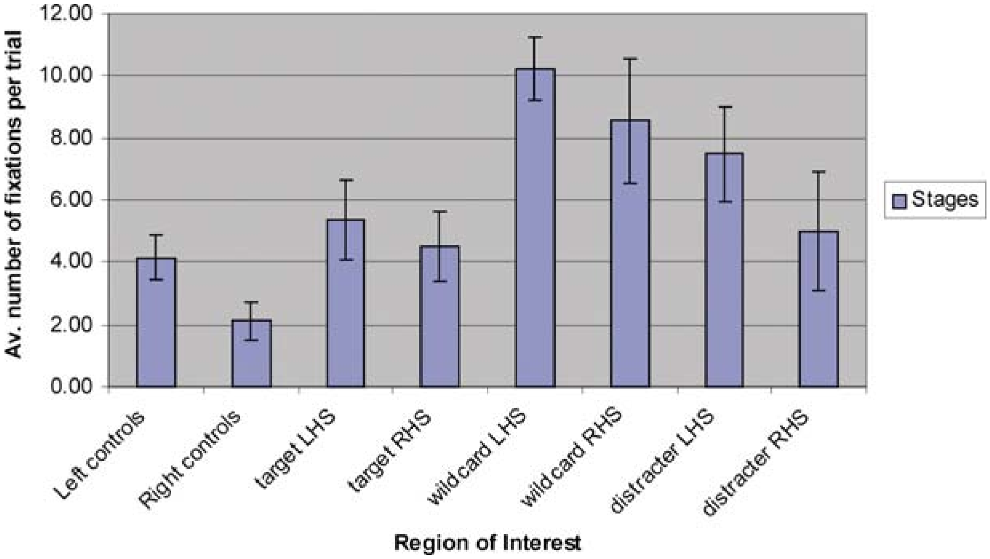


Fig. 4. Average number of fixations per trial on different ROIs, for each region of interest, for the Stages interaction technique. Error bars show 1 std. dev.

controls, the position of each fixation, and the dynamics of the eyegaze saccades. The pupil diameter was not used, as the ambient lighting was variable. The data was checked to see if there was any lost data due to blinks, head motion, etc. We only incurred minimal data loss, accounting for up to 50 ms per trial of lost data, as the subjects were in a chin rest.

We defined eight regions of interest: controls, targets, wild cards (potential targets), and distracters on each hemifield (recall that the left controls are for image navigation and the right controls are used to start and stop the study). We had to use visual inspection methods to count fixations on each region, as the centroid of the fixation clusters did not always land on a feature. An example is seen in Figure 2, fixation 6, which is likely to be on the target in the circle, but appears to be on the adjacent distracter. Furthermore, fixations on the controls were often just below the controls area, but above the objects in the stimuli, as seen for fixations 10 and 11 in Figure 2.

3. RESULTS

3.1 Response Time versus Interaction Technique

The average response time over all the subjects for the Stages interaction technique was 17.8 s/trial (SD 6.1) and for the FUI technique, 20.2 s/trial (SD 8.2). There is a significant difference between the Stages and FUI interaction techniques ($p = 0.026$ on a paired two sample for means over the four subjects).

3.2 Number of Fixations on Regions of Interest

The average number of eyegaze fixations per trial is 47.4 for Stages and 53.9 for the FUI interaction technique. Figure 4 shows a breakdown of the number of eyegaze fixations per trial on different ROIs for the Stages interaction technique. Excluding controls, there are many more fixations on the left hemifield images (23) than on the right hemifield images (18). The number of fixations on wild cards and distracters is proportional to their numbers in the stimuli images, so further results are presented only for the most important regions, the left and right controls and left and right targets.

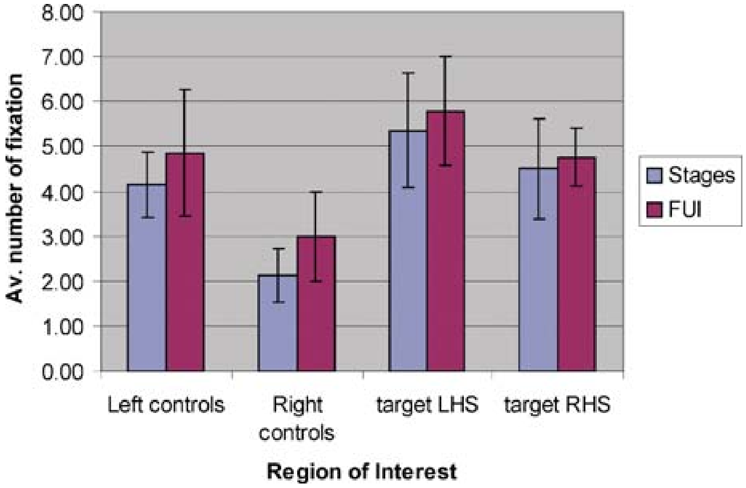


Fig. 5. Average number (SD) of fixations per trial on the controls and target in the right and left hemispheres, for each interaction technique. Error bars show 1 SD.

Table I. Duration of a Single Fixation (ms) per Trial (SD) and Cumulated Duration on Controls[a]

	Av. Fixation Dwell on Left Controls	Av. Fixation Dwell on Right Controls	Cumulated Duration on Left Controls	Cumulated Duration on Right Controls
Stages	314 (29)	371 (69)	1224 (309)	743 (309)
FUI	380 (56)	408 (37)	1924 (718)	1194 (566)

[a] Averaged over all the radiologists, for each interaction technique.

Figure 5 shows the average number of fixations per trial on controls and target ROIs under each interaction technique, averaged over all the radiologists.

Using a *t*-test paired two sample for means, there is not a significant difference between Stages and FUI for the number of fixations on the left controls ($p = 0.16$) or the number of fixations on the targets ($p = 0.36$ and $p = 0.41$ for LHS and RHS, respectively). However there is a significant difference on the right controls ($p = 0.046$).

3.2.1 *Duration of Fixations.* The average duration of a fixation on the left image elements (targets, wild cards, and distracters) for Stages and FUI are not significantly different ($p = 0.35$), averaging 288 ms over trials including both Stages and FUI on the left image, and 248 ms in the right image. However, the average duration of the fixation on the controls does show a difference between Stages and FUI. Table I shows the average duration of each fixation on the left and right controls and columns 3 and 4 show the cumulated dwell time per trial on the left and right controls, averaged over all the radiologists.

There is a significant difference between the two interaction techniques for the cumulated duration on the left controls ($p = 0.05$) and on the right controls ($p = 0.06$), with Stages being on average 36% faster than FUI on the left (image navigation) controls and Stages being 37% faster than FUI on the right controls.

3.3 Search Patterns and Strategies for Different Trial Outcomes

On starting a trial, eyegaze scan paths, such as seen in Figure 2, show most subjects follow a regular scan path until a target is observed, starting at the upper left side and searching more or less systematically down over the objects until a "suspicious object" is observed. First, there is an initial systematic search

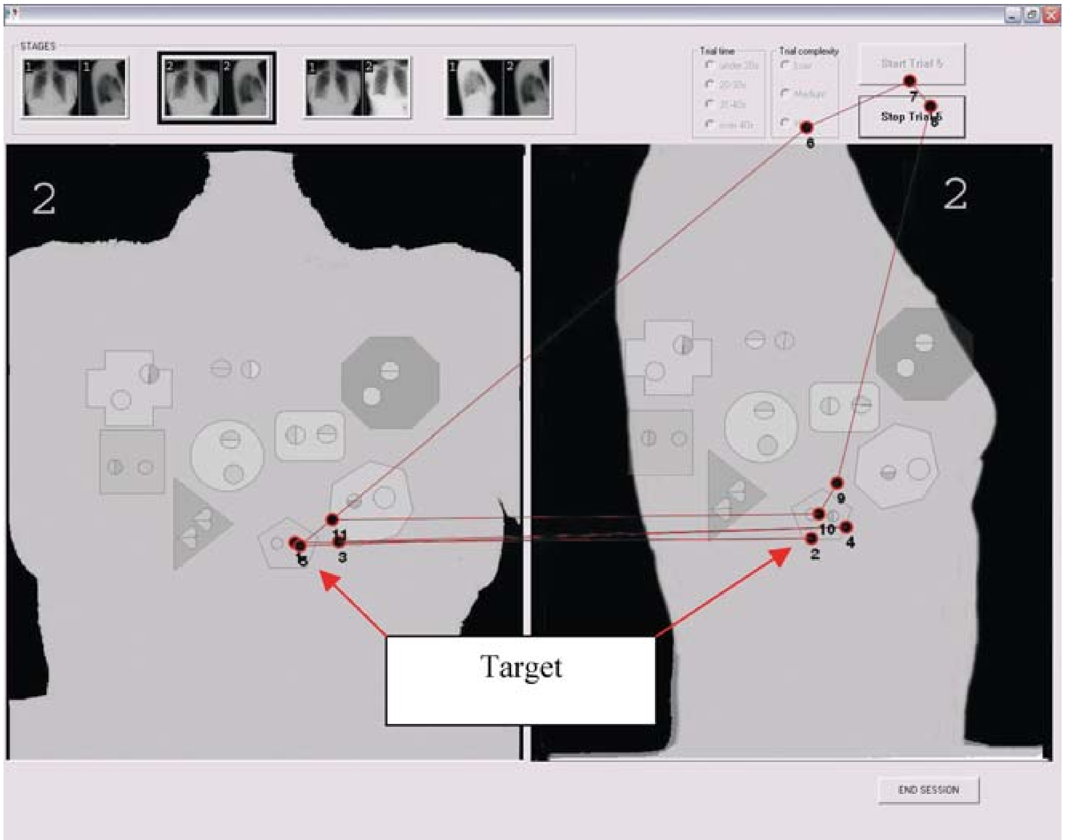


Fig. 6. Eyegaze pattern for study 2, where the target from study 1 is immediately fixated.

phase for the target. After the target has been recognized, there are then several transition fixations on the target during the recognition phase, before glancing up to the controls and stopping. The subject usually makes three to four fixations on the target in each hemifield to confirm or not the presence of a target. With an average fixation duration of around 268 ms, the cumulated fixations on the true target usually exceed 2000 ms. However, subjects make task-dependent short cuts. If a target was found in study 1, after opening study 2, the viewer can look straight at the position of the target from the first study, knowing that if there is a target in study 2, it will be in the same place. This situation is illustrated in Figure 6 where typically about 10 fixations are used in study 2.

The cumulated number of fixations versus trial outcome for study 1 and study 2 averaged over all the radiologists was calculated for each study. Fixations were least where a target was present in the first study, because an exhaustive search on study 2 was no longer required, as the subject knows there is at most one target in study 1 which may remain the same size, different size, or disappear in the second study (note the "short cut" is valid in this experiment, although this is not valid in real radiology tasks). When the target was present in both studies (outcome "11"), the radiologists occasionally performed additional comparisons to check for subtle changes in target size.

It was found that study 1 always had significantly more fixations than study 2, except for trials with outcome "00" in the FUI technique, where on average about 26 fixations were required to decide there was no target. However, for trials with outcome "00" in the Stages technique, on average 28 fixations were made in study 1 versus 23 fixations in study 2. This is likely because subjects can remember and fixate on the wild cards and not bother with the distracters.

3.4 Errors

Our subjects were instructed to be as accurate as possible, so a correct diagnosis is their primary task. Completing each trial in the shortest possible time interval was a secondary requirement. Thus, we made no references to the distribution of errors between the two interaction techniques: we traded time for accuracy. There were 17 errors in total; 8 errors with Stages and 9 errors with FUI, an insignificant difference. Of the 17 errors, 16 were false-negative errors where a target was missed and 1 was a false–positive error, where a distracter was taken as a target. No further analyses of these erroneous trials are presented here, as there is insufficient data for quantitative analysis.

4. DISCUSSION

4.1 Performance

With such a small sample, it is difficult to perform a definitive quantitative analysis on the differences between the two interaction techniques, although Stages is significantly faster than the FUI technique, with an average difference of 2400 ms per trial. From earlier work, we noted three other effects that contribute to the response time [Moise et al. 2005]: learning effects (Stages is easier to learn), image complexity, and trial outcome. Low-complexity images are faster than high-complexity images, as the search time increases with the number of potential targets in the image. The effect of trial outcome is discussed later. We use analysis of eyegaze fixations on the workstation controls to help determine why Stages is faster than FUI.

4.2 Eyegaze on Workstation Controls

The number of fixations on various ROIs for Stages is shown in Figure 4. From the average number of fixations/trial of 47.4 in Stages and 53.9 in FUI, we see that about 8.3% of the total number of eye-gaze fixations was spent on the left image navigation controls under the Stages condition, and 9% under the FUI condition. The percentage of fixations on the right controls was 4.2% and 5.5% for Stages and FUI respectively. Hence in total, 12.5% of the number of eye-gaze fixations was spent on the workstation controls under the Stages condition, and 14.5% under the FUI condition. Considering our controls are a simplified version of radiology workstation navigation controls, our results are consistent with previous eye-tracking research involving radiological interpretation of bone-trauma computed radiographs which reported a 20% distribution of eye-gaze over the workstation controls [Krupinski and Lund 1997].

The first column of Table I. shows that the average single fixation duration on left controls is 17% longer under FUI than under the Stages interaction technique. This leads to a significantly higher cumulated fixation duration on both the left and right controls for the FUI compared with the Stages interaction technique (columns 3 and 4 of Table I). This effect is expected, as the Stages technique was designed to reduce the image navigation effort; at best only two clicks are needed on the controls in Stages: one to move to the second study image pair, and one to stop the trial. However, there are often more than 6 fixations on the controls, explained by the eyegaze patterns like that in Figure 2, which show that several fixations may be made on the extended controls towards locating the next thumbnail.

Columns 3 and 4 of Table I show there is an extra average 700 msecs duration of fixations on the left (image navigation) controls for FUI over Stages, and 450 msecs on the right controls, compared with the increase in average response time of 2400 msecs. Explanation for these results is that the user has to make extra point-and-click movements to load the second pair of images in FUI. Using

Ruskin's low-level keystroke level model [Ruskin 2000], where the mouse actions can be broken down into constituent parts, a mouse move, or point (P), takes about 1.1 s, a keystroke (K) takes 0.2 secs and mental/perceptual processing (M) takes 1.35 secs. For a case with no target in either study, 2 point and 2 click movements in FUI are needed to select an image box and the iconic image to load into the box to see both images in study 2 whereas just one point and click is needed in Stages. Hence the control task takes an extra $(P + K) = 1.3$ secs in FUI, less than the observed difference of 2.4 secs. We see that the reduction in response time for the Stages technique is more than just the reduced time spent on the image navigation controls. In addition, the fixation duration on the right image controls is 450 msecs more in FUI than Stages, although these controls should only be used for starting and stopping the trials. This is likely due to a confusion in the subjects mind as to which controls they need to use to load new images in the FUI technique. Hence we hypothesise that the faster response times using Stages are due in part to the fact that the cognitive workload is reduced in Stages compared with the FUI, due to the disruption of the visual search process when viewing and manipulating the workstation controls.

4.3 Eyegaze Patterns

4.3.1 *Eyegaze on Left- and Right-Side Images.* Cognitive load increases fixation time [Findlay and Kapoula 1992]; hence, fixations are longer on areas of high information. Figure 4 shows there are significantly more fixations on the left image than on the right image (approximately in the ratio 4:3) and the fixations have longer duration on the left image than on the right image (288: 248 ms). This result is consistent with another example in mammography, where the left side "CC" image is fixated more than the right side "MLO" image [Kundel and Nodine 2004], although, in this situation, the CC was always presented on the left side, so it could not be determined whether radiologists preferred to look at the CC view or the left side. The reason for our task is that subjects tend to start at the left side and may use that image as a "baseline" image to compare individual features with the right-side image. Hence, the subjects tend to fixate their gaze for longer at the left-image than the right-image features, where the gaze may be made merely to confirm or not some visual feature identified in the left image. The average eyegaze fixation durations of 288 ms for left hemifield images and 248 ms for right hemifield images are consistent with other researchers' findings of mean fixation duration of 275 ms for visual search tasks [Rayner 1998].

4.3.2 *Eyegaze Search Patterns and Trial Outcome.* We have found that fixations occur in two phases, as others have reported [Pomplun et al. 2001b; Rayner 1998; Kundel 2004]. The first phase (*search phase*), occurs while the subject is searching the whole image for a suspicious object, or a likely target. The second (*recognition phase*), when fixations are used to confirm the target. Note that these two phases are readily observed in our comparative visual search task, because of the number of horizontal saccades across the hemifield images to compare the targets in each image during the recognition phase. This eyegaze pattern likely arises because our comparative visual search task for a target involves a match of several features, including the shape in which the target appears, the size of the target, the orientation of a dividing line within a target, and the location in the image. The human visual system has a limited short-term memory, for only four to five simple features [Ware 2004]. Therefore, to confirm each feature, horizontal saccades are performed, in preference to overloading the cognitive memory; usually at least three fixations on each side are made, and five horizontal cross-field saccades. Because of the complex nature of our targets, the limited capacity of the human visual memory, and the need for target comparisons between hemifield images, the cumulated dwell time on our true targets is often much higher than the 1000 ms reported by others.

4.4 Summary of Eyegaze Contributions

We have shown that the eyegaze data is useful to analyze computer interaction techniques, visual search strategies, and certain kinds of errors in our artificial look-alike radiology task. This may prove helpful in designing radiology workstations for real radiology tasks.

We found that eyegaze fixations on the controls did not account for all the increased time of the FUI technique over the Stages technique: disruption of the visual search may also affect response time.

Subjects take task-dependent "short cuts" to speed up search for targets, where the outcome affects the task time.

Although caution must be applied in such a small sample, we can assume that eyegaze analysis of our artificial look-alike radiology task can be effectively used to analyze workstation interaction techniques.

5. FUTURE WORK

We would like to perform more studies on more radiologists with more sophisticated interaction techniques to observe whether the eyegaze on the controls can be used to design better interaction techniques with reduced response times and faster learning. For example, it would be interesting to observe the impact of hanging protocols where the images are aligned vertically instead of horizontally.

We would like to study the effects of fatigue and experience on performance. Radiologists often work long hours and can become tired. The effect of a poor interaction technique may become more visible in the sense of more errors and longer response times when the subjects become more tired. As radiologists become more experienced, they could become faster if the workstation controls were properly designed, so the radiologists could preserve their gaze on images, instead of being distracted by manipulating controls. We would like to study the performance of radiologists before and after training, to see how the eyegaze pattern differs.

The next step is to evaluate radiologists performing real diagnosis tasks to observe in more detail whether eyegaze fixations and search paths can be used to highlight and possibly then eliminate false-negative decisions. Chest X-rays could be used as these are closest to our stimuli.

ACKNOWLEDGMENTS

Thanks to Dr. John Aldrich from Vancouver General Hospital, and the radiologists who performed the tasks. Funding was provided by the Canadian Natural Science and Engineering Research Council (NSERC) and the Advanced Systems Institute of British Columbia. Thanks also to the anonymous referees who helped us to focus and improve the paper.

REFERENCES

Ashman, C. J., Yu, J. S., and Wolfman, D. 2000. Satisfaction of search in osteoradiology. *Am. J. Roentgenol. 175*, 541–544.

ASL, Applied Science Laboratories. 2000. ASL Eye Tracker Manual, www.asl.com.

Berbaum, K. S., Franken, E. A, Dorfman, D. D., Miller, E. M., Caldwell, R. T., Kuehn, D. M., and Berbaum, M. L. 1998. Role of faulty decision making in the satisfaction of search effect in chest radiography. *Academic Radiol. 5*, 1, 9–19.

Berbaum, K. S., Brandser, E. A., Franken, E. A., Dorfman, D. D., Cadwell, R. T., and Krupinski, E. A. 2001. Gaze dwell time on acute trauma injuries missed because of satisfaction of search. *Academic Radiol. 8*, 4, 304–314.

Duchowski, A. T. 2000. A breadth-first survey of eye tracking applications. *Behavior Research Methods, Instruments and Computer 1*, 1–16.

Findlay, J. M. and Kapoula, Z. 1992. Scrutinization, spatial attention, and the spatial programming of saccadic eye movements. *Quart. J. Exp. Psycho.* 45A, 633–648.

Freedman, E. G. and Sparks, D. L. 2000. Coordination of the eyes and head: Movement kinematics. *Exp. Brain. Res. 131*, 22–32.

Hendee, W. R. 2002. The perception of images: I wouldn't have seen it if I hadn't believed it. In *Proc. SPIE Medical Imaging* 4686, Keynote Speech.

Hu C. H., Kundel, H. L., Nodine, C. F., Krupinski, E. A, and Toto, L. C. 1994. Searching for bone fractures: A comparison with pulmonary nodule search. *Academic Radiol. 1*, 25–32.

Humphrey, G. K. and Lupker, S. J. 1993. Codes and operations in picture matching. *Psychol. Res. 55*, 237–247.

Inamdar, I. and Pomplun, M. 2003. Comparative search reveals the tradeoff between eye movements and working memory use in visual tasks. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, Boston. 599–604.

Krupinski, E. A. 2000. The importance of perception research in medical imaging. *Radiat. Med. 18*, 6, 329–334.

Krupinski, E. A. and Lund, P. J. 1997. Differences in time to interpretation for evaluation of bone radiographs on monitors and film. *Academic Radiol. 4*, 177–182.

Krupinski, E. A., Kundel, H. L., Judy, P. F., and Nodine, C. F. 1998. The medical image perception society-key issues for image perception research. *Radiology 209*, 611–612.

Krupinski, E. A., Berger, W., Dallas, W., and Roehrig, H. 2004. Pulmonary nodule detection: What features attract attention? In *Proc. SPIE Medical Imaging* 5372. 122–127.

Kundel, H. L. 2004. Reader error, object recognition, and visual search. *Keynote speech*. In *Proc. SPIE Medical Imaging* 5372. 1–11.

Kundel, H. L. and Nodine, C. F. 2004. Modeling visual search during mammogram viewing. In *Proc. SPIE Medical Imaging* 5372. 110–115.

Kundel, H. L., Nodine, C. F., Thickman, D., and Toto L. 1987. Searching for lung nodules: A comparison of human performance with random and systematic scanning models. *Invest. Radiol. 22*, 417–422.

Kundel, H. L., Nodine, C. F., and Carmody, D. 1978. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Invest. Radiol. 13*, 3, 175–181.

Manning, D. and Ethell, S. C. 2002. Lesion conspicuity and AFROC performance in pulmonary nodule detection. In *Proc. SPIE Medical Imaging* 4686. 300–311.

Manning, D., Ethell, S. C., and Donovan, T. 2004. Categories of observer error from eye tracking and AFROC data. In *Proc. SPIE Medical Imaging* 5372. 90–99.

Mello-Thoms, C., Nodine, C. F., and Kundel, H. L. 2002. What attracts the eye to the location of missed and reported breast cancers? In *Proceedings of the IEEE Eye Tracking Research and Applications Conference, ETRA 2002*. 111–117.

Medical Image Perception Society, MIPS. 2005. http://www.radiology.arizona.edu/krupinski/mips/index.html.

Moise, A. 2003. *Designing better user interfaces for radiology workstations*, in Computing Science. PhD thesis, Simon Fraser University, Burnaby, V5A 1S3, Canada.

Moise, A. and Atkins, M. S. 2002. Workflow Oriented Hanging Protocols for Radiology Workstation. In *Proc. SPIE Medical Imaging* 4685. 189–199.

Moise, A. and Atkins, M. S. 2004a. Interaction techniques for radiology workstations: Impact on users' productivity. In *Proc. SPIE Medical Imaging* 5371. 16–22.

Moise, A. and Atkins, M. S. 2004b. Design Requirements for Radiology Workstations. *J. Digital Imaging 17*, 2, 92–99.

Moise, A. and Atkins, M. S. 2005. Designing better radiology workstations: Impact of two user interfaces on interpretation errors and user satisfaction. *J. Digital Imaging 18*, 2, 109–115.

Moise, A., Atkins, M. S., and Rohling, R. 2005. Evaluating different radiology workstation interaction techniques with radiologists and laypersons. *J. Digital Imaging 18*, 2, 116–130.

Nodine, C. F. and Kundel, H. L. 1987. Using eye movements to study visual search and to improve tumor detection. *Radiographics 7*, 1241–1250.

Nodine, C. F. and Kundel, H. L. 1990. A Visual Dwell Algorithm Can Aid Search and Recognition of Missed Lung Nodules in Chest Radiographs. In *Visual Search 2*. D. B. T. Francis, Ed. North-Holland, Amsterdam. 399–406.

Nodine, C. F., Kundel, H. L, Lauver, S. C., and Toto, L. C. 1996. The nature of expertise in searching mammograms for breast masses. In *Proc. SPIE Medical Imaging* 2712. 89–94.

Nodine, C. F., Kundel, H. L., Toto, L. C., and Krupinski, E. A. 1992. Recording and analyzing eye-position data using a microcomputer workstation. *Behavior Research Methods, Instruments, and Computers 24*, 475–485.

Nodine, C. F., Kundel, H. L., Mello-Thoms, C., and Weinstein, S. P. 2001a. Role of computer-assisted visual search in mammographic interpretation. In *Proc. SPIE Medical Imaging* 4324. 52–55.

Nodine, C. F., Mello-Thoms, C., Weinstein, S. P., Kundel, H. L., Conant, E. F., Heller-Savoy, R. E., Rowlings, S. E., and Birnbaum, J. A. 2001b. Blinded review of retrospectively visible unreported breast cancers: An eye-position analysis. *Radiology* 221, 122–129.

Pomplun, M. and Ritter, H. A. 1999. Three-level model of comparative visual search. In *Proceedings of the 21st Annual Meeting of the Cognitive Science Society*. 543–548.

POMPLUN, M., REINGOLD, E., AND SHEN, J. 2001a. Investigating the visual span in comparative search: The effects of task difficulty and divided attention. *Cognition* 81, B57–B67.

POMPLUN, M., SICHELSCHMIDT, L., WAGNER, K., CLERMONT, T., RICKHEIT, G., AND RITTER, H. 2001b. Comparative visual search: A difference that makes a difference. *Cognitive Science 25*, 1, 3–36.

RASKIN, J. 2000. *The Humane Interface*, Addison Wesley, Reading.

RAYNER, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychol. Bull. 124*, 3, 372–422.

ROGERS, L. F. 2000. Keep looking: satisfaction of search. *Am. J. Roentgenol.* 175, 287.

SALVUCCI, D. AND GOLDBERG, J. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of IEEE Eye Tracking Research and Applications Conference, ETRA 2000*. 71–78.

SCAR 2004. Society for Computer Applications in Radiology (SCAR) Newsletter, Fall 2004. SCAR Expert Hotline. *16*, 4.

WARE, C. 2004. *Information Visualization: Perception for Design*, 2nd ed. Morgan Kaufman.

YANG, G.-Z., DEMPERE-MARCO, L., HU, X.-P., AND ROWE, A. 2002. Visual Search: Psychophysical models and practical applications. *Image Vision Comput.* 20, 273–287.

Received January 2005; revised May 2005, June 2005; accepted June 2005